IEEE Xplore RELEASE 2.1

Welcome United States Patent and Trademark Office

Search Results BROWSE SEARCH IEEE XPLORE GUI

Results for "((((text categorization) and computer and training)<in>metadata)) <and> (pyr >= 1951 ..."
Your search matched **2** of **1351118** documents.
A maximum of **100** results are displayed, **25** to a page, sorted by **Relevance** in **Descending** order.

» Search Options

View Session History

New Search

Modify Search

((((text categorization) and computer and training)<in>metadata)) <and>

☐ Check to search only within this results set

Display Format: ◉ Citation ○ Citation & Abstract

view selected items **Select All** **Deselect All**

» Key

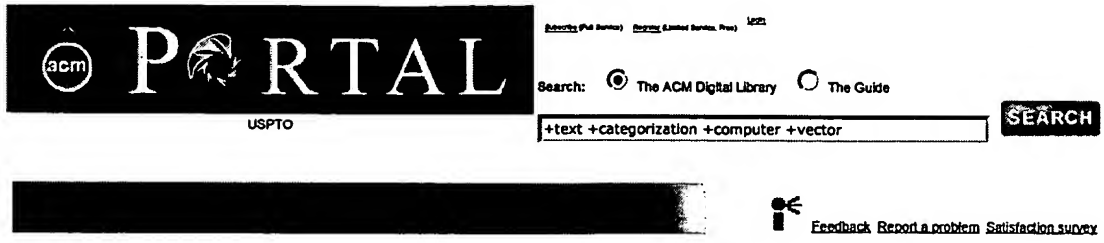
IEEE JNL	IEEE Journal or Magazine
IEE JNL	IEE Journal or Magazine
IEEE CNF	IEEE Conference Proceeding
IEE CNF	IEE Conference Proceeding
IEEE STD	IEEE Standard

☐ **1. Question classification for e-learning by artificial neur**
Ting Fei; Wei Jyh Heng; Kim Chuan Toh; Tian Qi;
Information, Communications and Signal Processing, 200; Pacific Rim Conference on Multimedia. Proceedings of the Conference of the Fourth International Conference on
Volume 3, 15-18 Dec. 2003 Page(s):1757 - 1761 vol.3
Digital Object Identifier 10.1109/ICICS.2003.1292768
AbstractPlus | Full Text: PDF(395 KB) **IEEE CNF**
Rights and Permissions

☐ **2. Text categorization using compression models**
Frank, E.; Chang Chui; Witten, I.H.;
Data Compression Conference, 2000. Proceedings. DCC .
28-30 March 2000 Page(s):555
Digital Object Identifier 10.1109/DCC.2000.838202
AbstractPlus | Full Text: PDF(48 KB) **IEEE CNF**
Rights and Permissions

Help

Indexed by Inspec®



Published before September 2003
Terms used text categorization computer vector

Found 865 of 144,021

Sort results by relevance

Display results expanded form

Save results to a Binder

Search Tips

Open results in a new window

Try an Advanced Search

Try this search in The ACM Guide

Results 1 - 20 of 200 Result page: **1** 2 3 4 5 6 7 8 9 10 next

Best 200 shown Relevance scale

1 A re-examination of text categorization methods

Yiming Yang, Xin Liu

August 1999 **Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(263.61 KB) Additional Information: full citation, references, citings, index terms

2 Special issue on special feature: Distributional word clusters vs. words for text categorization

Ron Bekkerman, Ran El-Yaniv, Naftali Tishby, Yoad Winter

March 2003 **The Journal of Machine Learning Research**, Volume 3

Publisher: MIT Press

Full text available: pdf(176.53 KB) Additional Information: full citation, abstract, index terms

We study an approach to text categorization that combines distributional clustering of words and a Support Vector Machine (SVM) classifier. This word-cluster representation is computed using the recently introduced *Information Bottleneck* method, which

generates a compact and efficient representation of documents. When combined with the classification power of the SVM, this method yields high performance in text categorization. This novel combination of SVM with word-cluster representation ...

3 Text categorization: A repetition based measure for verification of text collections and for text categorization

Dmitry V. Khmelev, William J. Teahan

July 2003 **Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval**

Publisher: ACM Press

Full text available: pdf(197.26 KB) Additional information: full citation, abstract, references, citings, index terms

We suggest a way for locating duplicates and plagiarisms in a text collection using an *R-measure*, which is the normalized sum of the lengths of all suffixes of the text repeated in other documents of the collection. The R-measure can be effectively computed using the suffix array data structure. Additionally, the computation procedure can be improved to locate the sets of duplicate or plagiarised documents. We applied the technique to several standard text collections and found that they ...

Keywords: cross-entropy, language modeling, text categorization, text compression

4 Meaningful term extraction and discriminative term selection in text categorization via unknown-word methodology

Yu-Sheng Lai, Chung-Hsien Wu

March 2002 **ACM Transactions on Asian Language Information Processing (TALIP)**, Volume 1 Issue 1

Publisher: ACM Press

Full text available: pdf(820.43 KB) Additional information: full citation, abstract, references, index terms

In this article, an approach based on unknown words is proposed for meaningful term extraction and discriminative term selection in text categorization. For meaningful term extraction, a phrase-like unit (PLU)-based likelihood ratio is proposed to estimate the

likelihood that a word sequence is an unknown word. On the other hand, a discriminative measure is proposed for term selection and is combined with the PLU-based likelihood ratio to determine the text category. We conducted several experim ...

Keywords: AC-machine, dimensionality reduction, discriminability, discriminative term selection, inconsistency problem, meaningful term extraction, n-gram, phrase-like unit, sparse data problem, term adaptation, term purification, text categorization, text indexing, unknown word detection, vector space modeling

5 Text categorization: A scalability analysis of classifiers in text categorization

Yiming Yang, Jian Zhang, Bryan Kisiel

July 2003 **Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval**

Publisher: ACM Press

Full text available: pdf(242.81 KB) Additional Information: full citation, abstract, references, index terms

Real-world applications of text categorization often require a system to deal with tens of thousands of categories defined over a large taxonomy. This paper addresses the problem with respect to a set of popular algorithms in text categorization, including Support Vector Machines, k-nearest neighbor, ridge regression, linear least square fit and logistic regression. By providing a formal analysis of the computational complexity of each classification method, followed by an investigation on the u ...

Keywords: complexity analysis, hierarchical text categorization, power law

6 Feature selection, perception learning, and a usability case study for text categorization

Hwee Tou Ng, Wei Boon Goh, Kok Leong Low

July 1997 **ACM SIGIR Forum , Proceedings of the 20th annual international ACM SIGIR conference on**

Research and development in information retrieval SIGIR '97, Volume 31 Issue SI

Publisher: ACM Press

Full text available: pdf(1.45 MB) Additional Information: full citation, references, citings, index terms

7 Special issue on special feature: A divisive information theoretic feature clustering algorithm for text classification

Inderjit S. Dhillon, Subramanyam Mallela, Rahul Kumar

March 2003 **The Journal of Machine Learning Research**, Volume 3

Publisher: MIT Press

Full text available: pdf(171.07 KB) Additional Information: full citation, abstract, citings, index terms

High dimensionality of text can be a deterrent in applying complex learners such as Support Vector Machines to the task of text classification. Feature clustering is a powerful alternative to feature selection for reducing the dimensionality of text data. In this paper we propose a new information-theoretic divisive algorithm for feature/word clustering and apply it to text classification. Existing techniques for such "distributional clustering" of words are agglomerative in nature and result in ...

8 Session: Exploring the use of linguistic features in domain and genre classification

Maria Wolters, Mathias Kirsten

June 1999 **Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics**

Publisher: Association for Computational Linguistics

Full text available: pdf(798.04 KB) Publisher Site Additional Information: full citation, abstract, references

The central questions are: How useful is information about part-of-speech frequency for text categorisation? Is it feasible to limit word features to content words for text classifications? This is examined for 5 domain and 4 genre classification tasks using LIMAS, the German equivalent of the Brown corpus. Because LIMAS is too heterogeneous, neither question can be answered reliably for any of the tasks. However, the results suggest that both questions have to be examined separately for each ta ...

9 Text categorization for multiple users based on semantic features from a machine-readable dictionary

Elizabeth D. Liddy, Woojin Paik, Edmund S. Yu

July 1994 **ACM Transactions on Information Systems (TOIS)**, Volume 12 Issue 3

Publisher: ACM Press

Full text available: pdf(1.17 MB) Additional Information: full citation, abstract, references, citings, index terms, review

The text categorization module described here provides a front-end filtering function for the larger DR-LINK text retrieval system [Liddy and Myaeing 1993]. The model evaluates a large incoming stream of documents to determine which documents are sufficiently similar to a profile at the broad subject level to warrant more refined representation and matching. To accomplish this task, each substantive word in a text is first categorized using a feature set based on the semantic Subject Field ...

Keywords: semantic vectors, subject field coding

10 Fast supervised dimensionality reduction algorithm with applications to document categorization & retrieval

George Karypis, Eui-Hong (Sam) Han

November 2000 **Proceedings of the ninth international conference on Information and knowledge management**

Publisher: ACM Press

Full text available: pdf(270.71 KB) Additional Information: full citation, references, citings, index terms

11 Special section on data mining for intrusion detection and threat analysis: Mining e-mail content for author identification forensics

O. de Vel, A. Anderson, M. Corney, G. Mohay

December 2001 **ACM SIGMOD Record**, Volume 30 Issue 4

Publisher: ACM Press

Full text available: pdf(954.21 KB) Additional Information: full citation, abstract, references, citings, index terms

We describe an investigation into e-mail content mining for author identification, or authorship attribution, for the purpose of forensic investigation. We focus our discussion on the ability to discriminate between authors

for the case of both aggregated e-mail topics as well as across different e-mail topics. An extended set of e-mail document features including structural characteristics and linguistic patterns were derived and, together with a Support Vector Machine learning algorithm, were ...

12 Classification: Boosting to correct inductive bias in text classification

Yan Liu, Yiming Yang, Jaime Carbonell

November 2002 **Proceedings of the eleventh international conference on Information and knowledge management**

Publisher: ACM Press

Full text available: pdf(199.02 KB) Additional information: full citation, abstract, references, index terms

This paper studies the effects of boosting in the context of different classification methods for text categorization, including Decision Trees, Naive Bayes, Support Vector Machines (SVMs) and a Rocchio-style classifier. We identify the inductive biases of each classifier and explore how boosting, as an error-driven resampling mechanism, reacts to those biases. Our experiments on the Reuters-21578 benchmark show that boosting is not effective in improving the performance of the base classifiers ...

Keywords: boosting, inductive bias, machine learning, text classification

13 Fast and accurate text classification via multiple linear discriminant projections

Soumen Chakrabarti, Shourya Roy, Mahesh V. Soundalgekar

August 2003 **The VLDB Journal — The International Journal on Very Large Data Bases**, Volume 12 Issue 2

Publisher: Springer-Verlag New York, Inc.

Full text available: pdf(456.36 KB) Additional information: full citation, abstract, citings, index terms

Abstract.Support vector machines (SVMs) have shown superb performance for text classification tasks. They are accurate, robust, and quick to apply to test instances. Their only potential drawback is their training time and memory requirement. For *n* training instances held in memory, the best-known SVM implementations

take time proportional to n^a, where *a* is typically between 1.8 and 2.1. SVMs have been trained on data sets with several thousand instances, but Web direct ...

Keywords: Discriminative learning, Linear discriminants, Text classification

14 Text classification using string kernels

Huma Lodhi, Craig Saunders, John Shawe-Taylor, Nello Cristianini, Chris Watkins

March 2002 **The Journal of Machine Learning Research**, Volume 2

Publisher: MIT Press

Full text available: pdf(1216.07 KB) Additional Information: full citation, abstract, references, citings, index terms

We propose a novel approach for categorizing text documents based on the use of a special kernel. The kernel is an inner product in the feature space generated by all subsequences of length k. A subsequence is any ordered sequence of k characters occurring in the text though not necessarily contiguously. The subsequences are weighted by an exponentially decaying factor of their full length in the text, hence emphasising those occurrences that are close t ...

Keywords: approximating kernels, kernels and support vector machines, string subsequence kernel, text classification

15 An example-based mapping method for text categorization and retrieval

Yiming Yang, Christopher G. Chute

July 1994 **ACM Transactions on Information Systems (TOIS)**, Volume 12 Issue 3

Publisher: ACM Press

Full text available: pdf(1.78 MB) Additional Information: full citation, abstract, references, citings, index terms

A unified model for text categorization and text retrieval is introduced. We use a training set of manually categorized documents to learn word-category associations, and use these associations to predict the categories of arbitrary documents. Similarly, we use a training set of queries and their related documents to

obtain empirical associations between query words and indexing terms of documents, and use these associations to predict the related documents of arbitrary queries. A Linear Le ...

Keywords: document categorization, query categorization, statistical learning of human decisions

16 Posters: Machine learning methods for Chinese web page categorization

Ji He, Ah-Hwee Tan, Chew-Lim Tan

October 2000 **Proceedings of the second workshop on Chinese language processing: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics - Volume 12**

Publisher: Association for Computational Linguistics

Full text available: pdf(706.21 KB) Additional Information: full citation, abstract, references

This paper reports our evaluation of *k* Nearest Neighbor (kNN), Support Vector Machines (SVM), and Adaptive Resonance Associative Map (ARAM) on Chinese web page classification. Benchmark experiments based on a Chinese web corpus showed that their predictive performance were roughly comparable although ARAM and kNN slightly outperformed SVM in small categories. In addition, inserting rules into ARAM helped to improve performance, especially for small well-defined categories.

17 Machine learning in automated text categorization

Fabrizio Sebastiani

March 2002 **ACM Computing Surveys (CSUR)**, Volume 34 Issue 1

Publisher: ACM Press

Full text available: pdf(524.41 KB) Additional Information: full citation, abstract, references, citings, index terms

The automated categorization (or classification) of texts into predefined categories has witnessed a booming interest in the last 10 years, due to the increased availability of documents in digital form and the ensuing need to organize them. In the research community the dominant approach to this problem is based on machine learning techniques: a general inductive process automatically builds a classifier by learning, from a set

of preclassified documents, the characteristics of the categories. ...

Keywords: Machine learning, text categorization, text classification

18 Hierarchical classification of Web content

Susan Dumais, Hao Chen

July 2000 **Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(1.16 MB) Additional information: full citation, abstract, references, citings, index terms

This paper explores the use of hierarchical structure for classifying a large, heterogeneous collection of web content. The hierarchical structure is initially used to train different second-level classifiers. In the hierarchical case, a model is learned to distinguish a second-level category from other categories within the same top level. In the flat non-hierarchical case, a model distinguishes a second-level category from all other second-level categories. Scoring rules can further take ad ...

Keywords: Web hierarchies, classification, hierarchical models, machine learning, support vector machines, text catergorization, text classification

19 Special issue on kernel methods: One-class svms for

document classification

Larry M. Manevitz, Malik Yousef

March 2002 **The Journal of Machine Learning Research**, Volume 2

Publisher: MIT Press

Full text available: pdf(203.03 KB) Additional Information: full citation, abstract, citings

We implemented versions of the SVM appropriate for *one-class* classification in the context of information retrieval. The experiments were conducted on the standard *Reuters* data set. For the SVM implementation we used both a version of Schoelkopf et al. and a somewhat different version of one-class SVM based on identifying "outlier" data as representative of the second-class. We report on experiments with different kernels for both of these implementations and with different represe ...

20 Text classification: Enhanced word clustering for hierarchical text classification

Inderjit S. Dhillon, Subramanyam Mallela, Rahul Kumar

July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**

Publisher: ACM Press

Full text available: pdf(993.07 KB) Additional Information: full citation, abstract, references, citings, index terms

In this paper we propose a new information-theoretic divisive algorithm for word clustering applied to text classification. In previous work, such "distributional clustering" of features has been found to achieve improvements over feature selection in terms of classification accuracy, especially at lower number of features [2, 28]. However the existing clustering techniques are agglomerative in nature and result in (i) sub-optimal word clusters and (ii) high computational cost. In order to expli ...

Results 1 - 20 of 200 Result page: **1** 2 3 4 5 6 7 8 9 10 next